AWARD NUMBER: W81XWH-12-1-0294

TITLE: High-Throughput Functional Validation of Progression Drivers in Lung Adenocarcinoma

PRINCIPAL INVESTIGATOR: Don L. Gibbons, M.D., Ph.D.

CONTRACTING ORGANIZATION: The University of Texas, MD Anderson Cancer Center
Houston, Tx 77030-4009

REPORT DATE: November 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
November 2014	Final	1 Sep 2012 – 31 Aug 2014

4. TITLE AND SUBTITLE
High-Throughput Functional Validation of Progression Drivers in Lung Adenocarcinoma

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0294

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Don L. Gibbons, M.D., Ph.D.

E-Mail: dlgibbon@mdanderson.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas, MD Anderson Cancer Center

1515 Holcombe Blvd.

Houston, TX 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The primary objective of our Early Investigator Synergistic Idea Award is to establish a driver prioritization pipeline to functionally evaluate lung cancer genomics data to identify somatic driver aberrations, which beyond the handful of well-characterized genes like oncogenic KRAS, contribute to lung cancer progression. As outlined in our FinalReport, we have successfully constructed the necessary gene libraries for the screens outlined in the proposal (Aim 1), optimized and conducted the proposed in vitro and in vivo screens (Aim 2). We have already identified several robust drivers of cell invasion and in vivo metastasis. After prioritizing our list of hits by filtering through clinical datasets, we have the necessary tools in place for further mechanistic studies of lead candidate genes (Aim 3), including our development of a novel isogenic human bronchial epithelial cell system that permits regulatible expression of the KRAS oncogene. We have made several technical improvements to the overall work flow of our functionalization pipeline and view this as a successful project that will facilitate further screening of emerging lung cancer genomics datasets extending well past the life of this particular grant.

15. SUBJECT TERMS
Lung cancer, driver oncogenes, high-throughput screens

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	12	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

INTRODUCTION

Cancer cells are endowed with diverse biological capabilities driven by an ensemble of inherited, somatic and epigenetic aberrations. As we enter the era of personalized medicine, characterization of the cancer genome has begun and will continue to influence diagnostic and therapeutic decisions in the clinic. Genome profiling technologies are generating a compendium of genomic aberrations in major cancer lineages with the goal of identifying the most promising therapeutic targets and diagnostic biomarkers. The output from these technologies is radically transforming cancer science. At the same time, these efforts are revealing the complexity of cancer genomes, which are comprised of causal "driver" aberrations and many more biologically neutral "passengers" that arise through the unstable nature of tumor genomes. While most cancers acquire one or more well-studied, high frequency driver events (e.g., mutations/gene copy number changes in *KRAS*, *TP53*, *EGFR*, *MYC*, *BRAF,* etc.), much less is known about the overly abundant low frequency (<5%) aberrations contributing to tumor progression and response to therapeutics. Comprehensive biological assessment of low frequency aberrations is difficult given their large number and the fact that they may either directly drive tumor progression or indirectly influence tumor behavior through modifying activities of other drivers like *KRAS*. Moreover, distinguishing driver events from passengers is further complicated by the fact that driver events are shaped by the specific biological context of a given cancer, including its tissue type, microenvironment and other host determinants including the immune system. The primary objective of the *Early Investigator Synergistic Idea Award* held by Dr. Kenneth Scott and myself is to establish a **driver prioritization pipeline** to functionally evaluate lung cancer genomics data to identify somatic driver aberrations, beyond the handful of well-characterized genes like oncogenic *KRAS* and *EGFR*, that contribute to lung cancer progression, invasion, and metastasis. Our study integrates genetically engineered mouse models of lung cancer, genomics data generated by The Cancer Genome Atlas (TCGA) and functional screens to identify drivers of lung cancer progression. We are using these tools to implement a novel, scalable screening infrastructure that permits high-content, gain-of-function screens to accelerate validation of functional somatic aberrations. This work is possible through advances made in our laboratories that include (1) high-throughput, highly accurate modeling of somatic aberrations into our collection of over 32,000 sequence-verified human genes and (2) a novel molecular barcoding approach that facilitates cost-effective detection of driver events following *in vitro* and *in vivo* functional screens. Our Specific Aims are as follows: (1) Construction of a lung cancer somatic driver library; (2) Functional screens for drivers of lung cancer metastasis; (3) Clinicopathological prioritization and validation of top candidates. **Our overall collaborative progress for the project is described in the report submitted by Dr. Kenneth Scott and herein I outline the specific contributions of the Gibbons lab in this collaborative "Synergy" grant.**

KEYWORDS

Metastasis, lung cancer, driver genes, high-throughput screen, *in vivo* screen

OVERALL PROJECT SUMMARY

Based on our approved Statement of Work, our overall goals, timeline and progress for the project were as follows:

Specific Aim 1: Construction of a lung cancer somatic driver library (proposed completion, months 3-6).

Please refer to the Year 1 overall Progress Report from Dr. Scott for this section, as the work was conducted in his laboratory. In brief, working with Dr. Chad Creighton at Baylor College of Medicine, a high-priority list of genes was selected by integration of multiple datasets from human lung tumors with genetically-engineered mouse models of lung adenocarcinoma from our group. A library of 260 candidates was constructed, with the ORFs for each gene cloned into the pLENTIEF6.3-puro lentiviral vector and each containing a unique molecular barcode for identification.

Specific Aim 2: Functional screens for drivers of lung cancer metastasis.

Subaim 2.1 *In vitro* screens for cell invasion and anoikis resistance (proposed completion, months 6-12)

In this Subaim we proposed to perform parallel *in vitro* screens for gene drivers of cell invasion and anoikis resistance using a 96-well format for cells transduced with single candidate genes from the ORF library constructed in Aim 1. As outlined in the first annual report, we began by performing extensive optimization of all steps required for these screens, including establishing cell plating and screening conditions for three separate cell lines: the optimization for the 393P and 393LN murine cell lines (described in our application) were performed in my laboratory and a new cell line with inducible mutant *KRAS* expression derived from human bronchial epithelial cells (HBECs) was performed by Dr. Scott's group.

Cell Invasion: Following the extensive optimization, the 393P cells were infected with lentivirus carrying the indicated genes, selected in puromycin for stable expression, and then screened for *in vitro* Transwell invasion using the established conditions and controls. Generation of the stable cell lines and cryo-banking were performed by Dr. Kundu in my lab and Dr. Wu in the Scott lab working together. Importantly, the decision was made to produce the cell lines in this fashion so that they could be cryo-banked for (1) future functional and biochemical validation assays and (2) *in vivo* screening assays proposed in Subaim 2.2 and as described below. Owing to this strategic decision, the cells were constructed in pools on a rolling basis and tested in both the *in vitro* and *in vivo* assays, which delayed the completion of the primary *in vitro* screen until the second year.

Completion of the 96-well Transwell invasion assays, as outlined in the Report from Dr. Scott, revealed increased invasion of 2.6-37-fold among the top 40 scoring gene candidates. This list included multiple genes reported to validate in our first Annual Report, e.g. *YWHAZ*, *MBIP* and *SRP54*. Thirty-one of the top 40 candidates were subsequently tested in my laboratory for secondary validation in the standard 24-well Transwell migration and invasion (containing Matrigel inserts) assays. Of the 19 additional genes tested in Year 2, 12 genes demonstrated increased migration and 11 demonstrated increased invasion (Figure 1A&B). Eight of these genes scored in both assays.

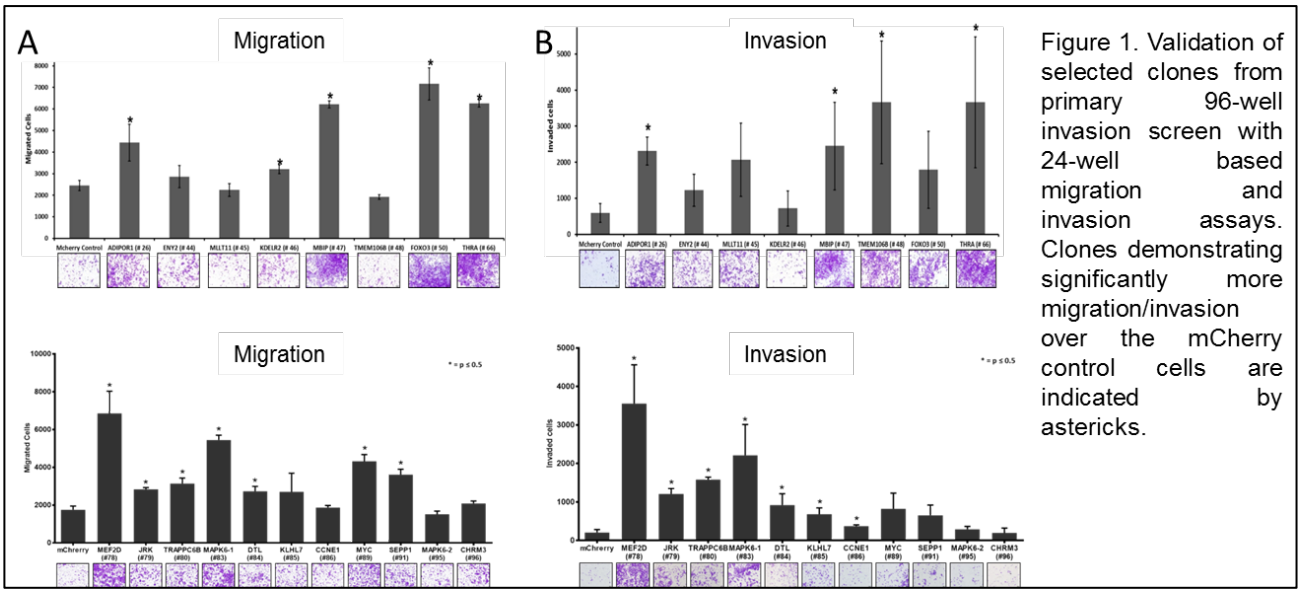

Figure 1. Validation of selected clones from primary 96-well invasion screen with 24-well based migration and invasion assays. Clones demonstrating significantly more migration/invasion over the mCherry control cells are indicated by astericks.

Anoikis Resistance: As described in our first annual report, we made great efforts to develop the proposed anoikis-resistance assays but were unsuccessful due to the fact that all cell lines assayed (e.g., 393P, 393LN, etc) exhibited robust anoikis resistance and were therefore not suitable as screening models. During Year 2 the newly-derived HBEC cell model (see first Annual Report) was assessed in Dr. Scott's lab for its performance in this assay, but unfortunately, this line also exhibited high baseline anoikis resistance rendering it unsuitable for screening. In our first annual report we

described an alternative 3D culture assay, to be used in place of the anoikis assay if problems persisted, whereby cells are grown on a bed of extracellular matrix. As we have previously published, this method better mimics the *in vivo* condition by modeling the cell-matrix interactions, is scorable for cell growth, cell-cell organization and invasion, and can be modified by adjusting the composition and biophysical properties of the matrix [1-3]. Using this secondary screen, Dr. Kundu in my laboratory

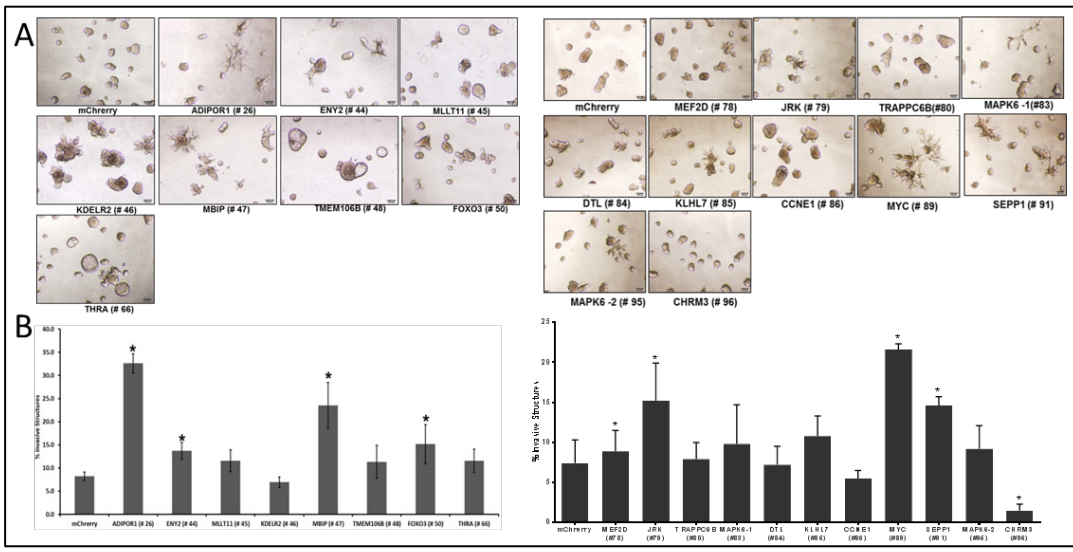

Figure 2. 3D *in vitro* growth and invasion assays. Single cell suspensions of the indicated stable 393P clone-expressing cell lines were plated onto beds of Matrigel/type 1 Collagen. After growth for one week, colonies were scored for size and percent invasion vs. mCherry expressing negative controls. Clones with significantly more invasion are indicated by astericks.

validated several of the clones identified during the first year of work (see first annual report), and during the second year tested 19 clones identified from the primary *in vitro* screen. These results demonstrated significantly more invasive structures in 8 of the clones compared to Mcherry control cells (Figure 2A & B). Four of these genes also scored in **both** the Transwell migration and invasion assays described above, while 3 of the remaining scored in **either** the Transwell migration or invasion assays. These data confirm the robustness of the primary *in vitro* screen and begin to indicate potential underlying mechanisms that drive the phenotype for each of the genes that scored, with some primarily enhancing motility under all conditions while others enhance cell-matrix interactions. Additionally, several of the genes appear to drive hyperproliferation and disorganized growth, but not frank invasion into the extracellular matrix.

Subaim 2.2. *In vivo* positive selection screens (proposed completion months 3-15).

In this Subaim we proposed to perform *in vivo* metastasis screens with pooled viral-infected cells to positively select for single and combinatorial drivers of metastasis. Pilot assays outlined in the first annual report demonstrated that subcutaneous rather than orthotopic injection of 10^6 cells provided an optimal period of primary tumor growth that could be monitored over ~6-8 weeks before the animals required sacrifice due to primary tumor burden, thus allowing sufficient time for metastasis and detection of metastasis drivers. Since the goal of the *in vivo* screen was to recover distant metastases, the longer period of primary tumor growth was favored. This approach also facilitated the monitoring of primary tumor growth rate as an independent phenotypic readout. During the second year of work, regulatory re-approval of the animal protocols from the MD Anderson IACUC and the DoD IACUC was required to continue the *in vivo* screen. With approval, we used the cells infected and selected for stable ORF expression to conduct the *in vivo* screen. Each group included 10 syngeneic immunocompetent 129Sv mice that received a pool of 10^6 cells each, comprised of ORF expressing cell lines (50,000 cells/ORF) plus one additional mCherry control cell line as an internal control. Mice were observed for general health and the growth of primary subcutaneous flank tumors recorded. The animals were sacrificed when the primary tumor burden was excessive or the mice showed apparent deterioration in general health. This work has all been performed in the Gibbons lab and in the MDACC animal facility. Not all of the candidate genes produced stable cell lines, so we

Pool #	Clones Injected	Mice Injected	Mice Euthanized	Mice alive with tumors	Mice with Metastases	Mice with no metastases	Identified metastasis drivers (by sequencing)	Identified tumor drivers (by sequencing)
1	19+Mcherry	10	10	0	2	8	TBD	TBD
2	19+Mcherry	10	10	0	2	8	TBD	TBD
3	19+Mcherry	10	10	0	8	2	9	4
4	19+Mcherry	10	10	0	8	2	3	2
5	19+Mcherry	10	10	0	3	7	4	4
6	19+Mcherry	10	10	0	2	8	2	3
7	19+Mcherry	10	10	0	9	1	TBD	TBD
8	19+Mcherry	10	9	1	6	3	TBD	TBD
9	19+Mcherry	10	10	1	8	2	TBD	TBD
10	19+Mcherry	10	10	0	6	4	TBD	TBD
11	19+Mcherry	10	10	0	9	1	TBD	TBD
12	10+Mcherry	10	4	6	3	1	TBD	TBD
Total	218	120	113	7	66	47	18	13

Table 1. Summary of data from the primary *in vivo* screen. Pink highlight denotes pools that have been sequenced, blue denotes pools for which sequencing data is pending, and the remainder are being processed for sequencing (TBD, to be determined).

have entered 219 of the 260 candidates into *in vivo* screens across 12 separate cohorts, including 120 mice (Table 1).

Of the 120 mice which received the injected pooled clones, all have demonstrated robust subcutaneous tumor formation. At euthanasia the animals receive a full necropsy to evaluate primary tumor size/weight, evidence of distant metastases, and for collection of tumor tissue (Figure 3 & 4). As outlined in the report from Dr. Scott, the driver pro-viral inserts are identified by their associated barcodes. Genomic DNA is purified from collected tumor tissues, barcodes are amplified by PCR and the barcode products sequenced and quantitated by an Ion Torrent Personal Genome Machine (PGM), which can multiplex up to 96 tumor samples, thus significantly reducing overall cost. Enrichment is defined as candidates that are significantly higher in output (tumor or metastases) than input (injected cells). We observe that, relative to injected cells, tumors and metastases **positively select** driver genes and lose those with no role in tumor growth or metastasis (i.e., passengers). Of the 113 mice that were euthanized so far across the different pools, we observed 66 mice with varying numbers of lung metastases. Seven mice are still undergoing observation. From 21 mice that showed lung metastases (Pools 3-6), we extracted genomic DNA from primary tumors and metastatic lung nodules. We generated barcode libraries and sequenced these by Ion Torrent sequencing, to identify 4 genes that were enriched in primary tumors only, 9 metastasis drivers and 9 that were enriched in primary tumors and metastases (Table 2 & Figure 4). As shown in the Table, four of the identified *in vivo* hits were also in the Top 10 or Top 40 hits from the primary *in vitro* invasion screen. This suggests that the parallel approach found reasonable overlap, but also yielded complementary genes that would have been missed with either approach alone. Currently we are processing tumor and lung tissues from 16 additional mice which showed lung metastasis for genomic DNA extraction and barcode sequencing to identify additional potential *in vivo* metastasis and tumor growth drivers.

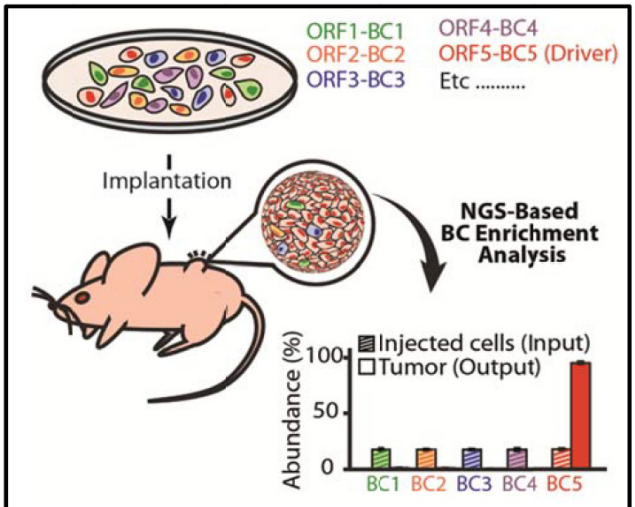

Figure 3. Barcode enrichment analysis. See text for details.

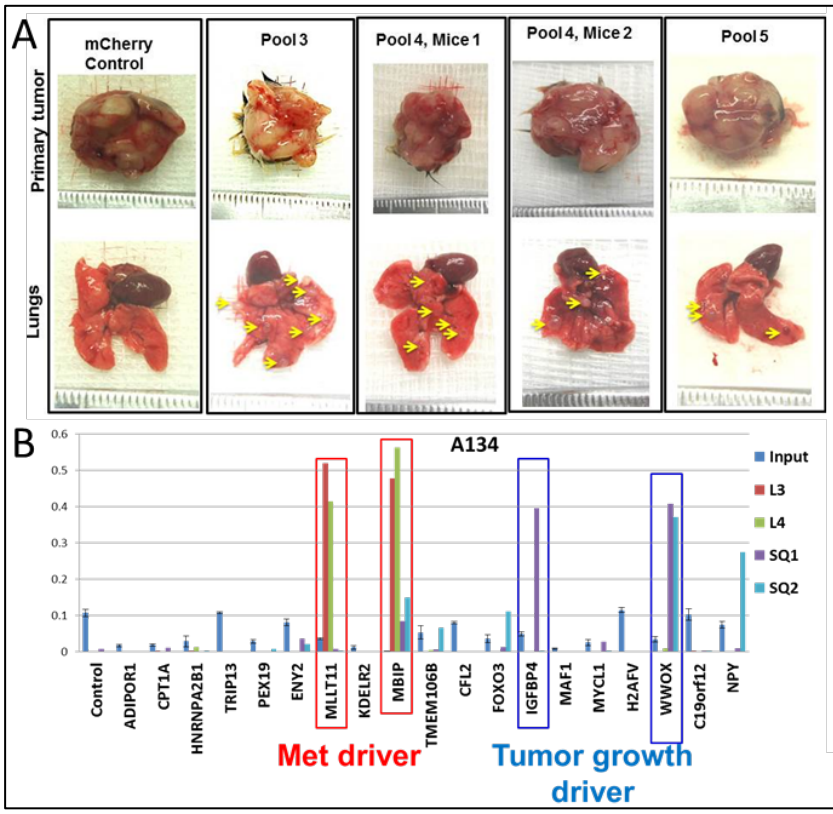

Figure 4. A) Representative images of the primary tumor and lungs with metastatic nodules (arrows) from mice injected with different pools of candidate cell lines. **B)** Candidate barcodes were detected by Ion Torrent sequencing of genomic DNA extracted from either the pooled input cells (input), sections of the primary tumors (SQ1/SQ2/SQ4) or macro-dissected lung metastatic nodules (L3/L4/L5). Plots represent fractional barcode sequence reads for respective pooled candidates from various tissue as indicated, for representative mouse injected with different pools of candidate cell lines. Candidates indicated in red boxes are the identified potential metastasis drivers due to enrichment only in the lung metastatic nodules and not in the input cells or the SQ tumors. Candidates indicated in blue boxes are the identified potential tumor growth drivers as they are differentially enriched in the SQ tumors only.

To validate the documented *in vivo* hits identified by sequencing, we have taken 4 that emerged (TMEM106b, NPY, C19ORF12, and GNAS), regenerated the cell lines by infection of the ORF-containing lentivirus and selection, and re-injected them into syngeneic animals (10 animals/clone) individually with only mCherry expressing cells as an internal control to test their primary tumor growth and ability to induce lung metastasis. The experiments are still on-going, but 5 of 15 animals already analyzed demonstrate distant metastases for 3 of the 4 genes, confirming the initial observations from the pooled *in vivo* screen. We have also taken these same four *in vivo* hits and begun *in vitro* investigations to evaluate their phenotypic behavior and identify the potential mechanism(s) driving metastasis. Interestingly, none of

LIST OF *IN VIVO* HITS

						in vitro	
Gene	Total mice injected	Total mice with mets	Total mice with gene enrichment	% mice/total injected	Enrichment (SubQ/Met)	Top 10	Top 40
TMEM106B	10	6	5	50%	Tumor & Met Driver	no	no
WWOX	10	8	4	40%	Tumor & Met Driver	no	no
NPY	10	8	4	40%	Tumor & Met Driver	no	no
MYC	10	3	3	30%	Metastasis Driver	yes	
C19ORF12	10	8	2	20%	Metastasis Driver	no	no
CCNE1	10	3	2	20%	Tumor & Met Driver	yes	
STK3	10	2	2	20%	Tumor & Met Driver	no	no
GNAS	10	6	2	20%	Metastasis Driver	no	no
KRTCAP2	10	3	2	20%	Tumor Growth Driver	no	no
MBIP	10	8	1	10%	Metastasis Driver		yes
FMOD	10	2	1	10%	Tumor Growth Driver		yes
MLLT11	10	8	1	10%	Metastasis Driver	no	no
HNRNPA2B1	10	8	1	10%	Metastasis Driver	no	no
H2AFV	10	8	1	10%	Metastasis Driver	no	no
TRIP13	10	8	1	10%	Tumor & Met Driver	no	no
PEX19	10	8	1	10%	Tumor & Met Driver	no	no
RASSF8	10	6	1	10%	Metastasis Driver	no	no
WIPI2	10	6	1	10%	Tumor Growth Driver	no	no
STY11	10	3	1	10%	Tumor Growth Driver	no	no
RALA	10	3	1	10%	Tumor & Met Driver	no	no
LACTB2	10	3	1	10%	Metastasis Driver	no	no
CA13	10	2	1	10%	Tumor & Met Driver	no	no

Table 2. Candidate drivers from pooled primary *in vivo* screen identified by sequencing.

the four have enhanced migration or invasion in standard Transwell assays, but were significantly more invasive in 3D growth on matrix (Figure 5 & 6). Similar secondary *in vivo* validations and *in vitro* testing for additional candidates identified from the primary screen are planned, but are awaiting the pending additional sequencing results from the primary *in vivo* screen.

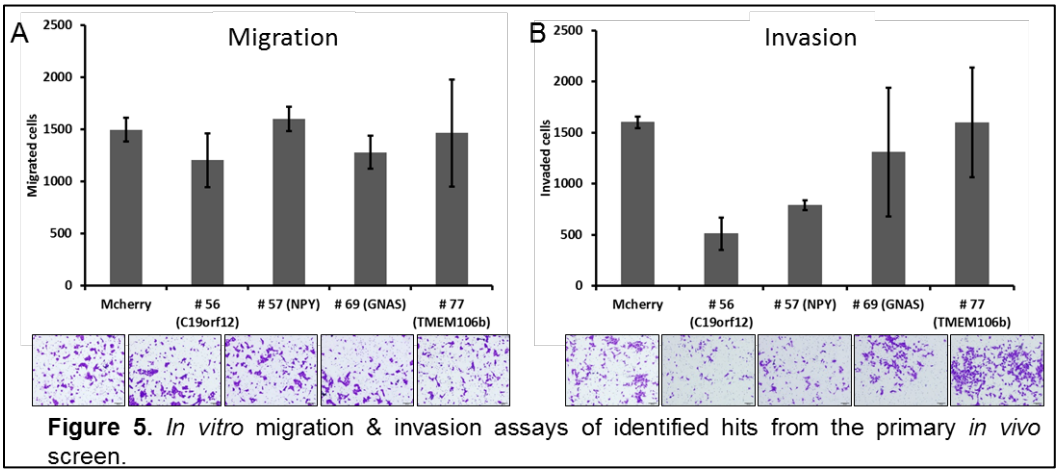

Figure 5. *In vitro* migration & invasion assays of identified hits from the primary *in vivo* screen.

***In vivo* validation of *in vitro* hits**. We have also taken individual genes that scored in the primary *in vitro* screens and started the *in vivo* validation studies to test their ability to induce lung metastasis by injecting them individually into syngeneic mice. Two of the genes reported in our first annual report, YWHAZ and SRP54, were tested and are capable of driving *in vivo* metastases in 60% and 80% of the animals tested, respectively. These hits will be carried forward into the investigation of potential mechanistic pathways that these drivers genes regulate that induce metastasis.

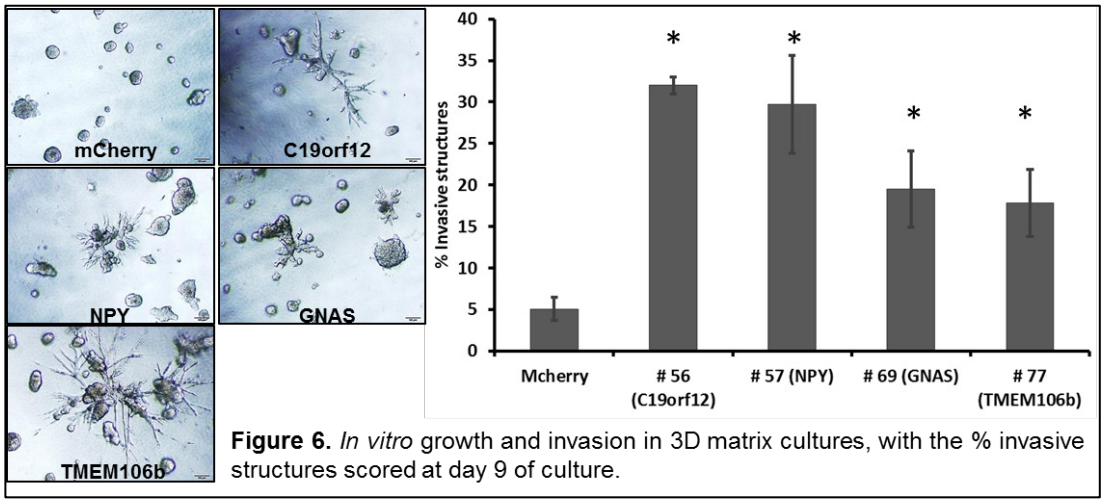

Figure 6. *In vitro* growth and invasion in 3D matrix cultures, with the % invasive structures scored at day 9 of culture.

Specific Aim 3: Clinicopathological prioritization and validation of top candidates.

In this Aim we proposed to take forward identified drivers from the *in vitro* and *in vivo* screens for clinical validation and prioritization by analysis of their expression in publically available datasets (e.g., TCGA and others, as well as in-house datasets with clinical annotation). We also proposed to begin functional and mechanistic studies of clinically validated high priority candidates.

Subaim 3.1. Clinical validation and prioritization of metastasis genes (proposed completion during Year 2).

Working with Dr. Creighton, we analyzed the top 40 *in vitro* hits for clinical relevance in a compendium of multiple public databases that encompasses expression data for ~1500 lung tumors and copy number alterations for the 230 tumors in the TCGA dataset for lung adenocarcinoma. From this analysis we identified 10 genes (Table 3) that demonstrate increased *in vitro* invasive phenotype that also significantly correlated with poor survival and high frequency of amplification in TCGA LUAD

samples (or for two of the genes, mutation rather than amplification). These genes have already undergone the secondary *in vitro* validations outlined in Subaim 2.1 and have been prioritized for further mechanistic study. Some of the candidate genes have been independently identified in the *in vivo* metastasis screen and others were part of the *in vivo* pools for which sequencing data is pending. Once the sequencing of *in vivo* candidates is finished, a similar analysis will be performed for the top-scoring hits. For those that are validated, additional evaluation of patient samples by immunohistochemistry (IHC) will be conducted in collaboration with Dr. Ignacio Wistuba's group, who has samples from early-stage surgical resections (similar to the TCGA dataset) that have had expression profiling and for which tissue exists for IHC. He also has similar datasets and samples from late-stage metastatic specimens, which will allow us to compare expression of the identified markers with treatment outcome. In fact, *MBIP* is one of the genes identified from the *in vivo* screen, validated in our *in vitro* experiments and which Dr. Wistuba's group independently identified as part of a 12-gene prediction signature of outcome in surgically resected cases [4].

RANK	Symbol	Entrez ID	Survival correlation (compendium n=1492) beta (>0, corrlation with worse outcome)	Survival correlation (compendium n=1492) p-value	% gain (3+ copies) TCGA (n=493)
1	MYC	4609	0.098748	0.016225	14.0%
2	YWHAZ	7534	0.111509	0.005496	10.5%
3	KDELR2	11014	0.103565	0.010539	7.7%
4	DFNA5	1687	0.138023	0.000877	7.3%
5	IMPAD1	54928	0.09872	0.020245	6.7%
6	SNAI2	6591	0.093286	0.023168	6.1%
7	DTL	51514	0.233929	4.78E-08	5.5%
8	CCNE1	898	0.141911	0.000241	4.5%
9	CBLB*	868	0.088768	0.031021	1.0%
10	MAPK6*	5597	0.191224	2.07E-06	0.2%

Table 3. Top 10 genes identified to be highly invasive *in vitro* and were found to correlate with poor survival and high frequency of amplification in the TCGA LUAD samples. ***** These two genes are not amplified, but frequently found to be mutated.

Subaim 3.2. Functional and mechanistic study of the lead metastasis genes (proposed completion during Year 2).

Owing to the delay in completion of the primary *in vitro* invasion screen and the on-going sequencing of metastases from the *in vivo* screen, completion of this part of aim 3 has subsequently been delayed. However, we are currently entering the top hits into the engineered HBEC system to evaluate their dependency on mutant *KRAS* (as described in more detail in the report from Dr. Scott), especially since bioinformatic analysis by Dr. Creighton demonstrated that several of the genes are expressed at disproportionately higher frequencies in tumors with mutant *KRAS*. For oncogenic and metastasis driver validation assays, we proposed to use cancer cell lines for appropriate expression or knock-down studies. Depending upon the particular genes identified in the screens and the observed phenotypes, we will use human NSCLC cancer cell lines or other cancer cell types as appropriate.

Our initial characterization has already revealed that the genes stratify into those that only appear to drive metastasis, versus those that enhance primary tumor growth and metastasis. Additionally, based upon *in vitro* growth characteristics, some of the identified drivers appear to induce an EMT-like phenotype. Certainly one of the top hits (SNAI2) is a well-characterized inducer of EMT, which is consistent with our prior publications on the ability of EMT inducers to drive progression and metastasis in this murine lung cancer model. Additional work will be required to more fully

characterize the effects of the candidate genes. Some of this work has already begun, but will be beyond the scope of this current grant. This work will be carried forward once we have further confirmation of secondary validation from *in vivo* experiments, which may help guide the investigations, especially if multiple genes point to a common biologic pathway.

KEY RESEARCH ACCOMPLISHMENTS

- **Validation of the high-throughput *in vitro* invasion screens with standard larger format transwell assays:** Allows the rejection of false positives from the primary screens and identifies potential genes for further in vitro testing.

- **Robust 3D invasion assay:** Better mimics the 3D nature and extracellular matrix components found *in vivo*, while allowing manipulation of the matrix and easy scoring of the cellular phenotype.

- **Positive *in vivo* growth and metastasis screen:** Ability to directly test the *in vivo* role of genes on primary tumor growth and metastasis in a medium-throughput, pooled fashion due to the innovative barcoding techniques, use of sensitive sequencing and the positive selection of the *in vivo* screen.

- **Validation of candidates from the parallel *in vivo* and *in vitro* screens:** Allows for confirmation of hits from the primary *in vivo* screen, exclusion of any false positives and reconsideration of any candidates that did not score *in vivo* but that produced strong activity in the primary *in vitro* screen.

CONCLUSION

Our overall goal with this project is to establish a pipeline of robust screening techniques to functionally prioritize the data emerging from large-scale genomics efforts in lung cancer. Using a combination of *in vitro* and *in vivo* screens we have been able to identify and validate oncogene and metastasis drivers, and generate the pre-clinical cell and animal models needed for therapeutic targeting studies. As outlined in the Body of the main report from Dr. Scott, we have successfully constructed the necessary libraries for this work, conducted both the *in vitro* and *in vivo* screens, performed the bioinformatic analyses to select for top candidates to carry forward into mechanistic studies. We were able to incorporate several technical improvements to our original work plan to improve the workflow and sensitivity of both the in vitro and in vivo screens. In addition to the multiple candidates that have been identified and require additional mechanistic work, the pipeline is now established to extend the screening efforts past the life of this particular grant. We also feel that these techniques can be broadly implemented for functionalization of genomic data for other tumor types, e.g. pancreatic cancer.

PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS.

High-throughput functional screen for metastasis drivers in lung cancer.
Samrat T. Kundu, Ping Wu, Chad J. Creighton, Kenneth L. Scott, and Don L. Gibbons, Cold Spring Harbor Laboratory Meeting, Mechanisms & Models of Cancer, August 12-16, 2014.

INVENTIONS, PATENTS AND LICENSES.

None.

REPORTABLE OUTCOMES.

- **Pro-invasion genes:** Primary screens and secondary validation have revealed 39 genes genes that robustly drive cell invasion.

- **Pro-metastasis genes:** Primary screens have already identified 18 metastasis drivers, many not previously linked to lung cancer progression. Sequencing data is still pending for additional *in vivo* experiments, so this list will likely grow by a few.

- **Genomic functionalization screening pipeline methodology:** Overall, this project has produced a robust methodology to create stable clones expressing wild-type or mutant ORFs based upon selected gene lists from human cancer genomic profiles, enter them into parallel *in vitro* and *in vivo* assays to phenotypically test for drivers of growth, migration, invasion and metastasis. This will be useful for continued characterization of the available lung cancer datasets, but is also potentially applicable to other epithelial tumor types.

OTHER ACHIEVEMENTS.

Based upon this work, Dr. Kundu has applied for a one-year Career Development Award through our MDACC/UT Southwestern Lung Cancer SPORE.

Based partly upon this work, Dr. Scott obtained a CPRIT HR/HI award from the State of Texas to extend this conceptual screening methodology into GEMMs.

REFERENCES.

1. Ahn, Y.H., et al., *ZEB1 drives prometastatic actin cytoskeletal remodeling by downregulating miR-34a expression.* J Clin Invest, 2012. **122**(9): p. 3170-83.
2. Gibbons, D.L., et al., *Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression.* Genes Dev, 2009. **23**(18): p. 2140-51.
3. Gill, B.J., et al., *A synthetic matrix with independently tunable biochemistry and mechanical properties to study epithelial morphogenesis and EMT in a lung adenocarcinoma model.* Cancer Res, 2012. **72**(22): p. 6013-23.
4. Tang, H., et al., *A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients.* Clin Cancer Res, 2013. **19**(6): p. 1577-86.

APPENDICES.

None